AI 时代出版产业融合发展的策略研究

——以基于“CHESS 战略”开展 AI 有声书业务为例

Research on the Strategy of Integrated Development of Publishing Industry in AI Era

--An Example of AI Audiobook Business Based on "CHESS Strategy"

张　聪 1）　　丁彦清 2）

1）北京印刷学院，102627，北京；2）北京电影学院，100091，北京

摘　要：人工智能技术发展迅速，出版产业在“四全媒体”的指导下积极融合新技术，AI 语音技术（TTS、ASR）是落地最早、应用最广的人工智能方向。本文在“CHESS 战略”的理论框架下研究出版社发展基于 AI 语音技术的有声书业务的策略，首先以“媒体可供性”理论阐释出版业发展 AI 有声书的必要性，再以技术、政策、市场、人才、管理、资产六个维度的产业驱力来分析出版社参与 AI 有声书产业面临的困境，最后提出对于出版社开展 AI 有声书业务的建议与展望。

Artificial intelligence is developing rapidly, and the publishing industry is actively integrating new technologies under the guidance of "Four Dimension Media", and AI speech technology (TTS, ASR) is the earliest and most widely used AI direction. In this paper, we research the strategy of publishing houses to develop audiobook business based on AI speech technology under the theoretical framework of "CHESS Strategy", firstly, we explain the necessity of developing AI audiobook in the publishing industry with the theory of "Media Affordance Theory". We then analyze the dilemmas faced by publishers participating in the AI audiobook industry in six dimensions: technology, policy, market,

manpower, management, and capital, and finally propose suggestions and prospects for publishers to develop the AI audiobook business.

关键词：有声书；AI 语音技术；媒体可供性；产业融合理论

近年来我国 AI 语音技术进入应用快速落地期，相较 2016 年之前在情感语音合成与自然语义理解等方向取得显著突破。国内外相关企业都纷纷开放语音生态系统，以“产业融合”的方式将 AI 语音技术应用于各种场景，并且在短视频创作、虚拟偶像（主播）、智能客服、智慧教育、智能汽车等行业取得成功的商业应用。AI 语音技术的两大分支：语音合成（TTS：Text To Speech“从文本到语音”）、语音识别（ASR：Automatic Speech Recognition“从语音到文本”）的技术原理使其天然地适合应用于以文字编辑为工作的出版行业。在快速发展的“声音经济”中，出版社可以通过应用 AI 语音技术解决音频内容制作能力弱的问题，以极低的成本快速生成人耳难辨真伪的音频内容，使其在出版、网络音视频、AI 的“产业融合”中获得更大发展空间。

一、以“媒体可供性”理论阐释出版业发展 AI 有声书的必要性

“可供性”的概念源自心理学领域，原意指生物(或行为主体) 在物理环境中潜在的各种行动的可能性，它源自主体对效用的主观感知与技术的客观品质之间的相互作用。[1]后由传播学学者潘忠党引用于新媒体研究，现已成为理解新的媒介现象，评估新的媒介技术、形态、结构发展潜力的一种系统的媒介研究框架。[2]媒体可供性强调技术对传播实践的塑造，认为在新技术改造媒介过程中为受众带来的新“可供性”，不是媒体创新预设的目标而是在传播实践下技术、媒体、受众相互作用的结果。因此“媒体可供性”理论框架为传统媒体的融合创新提供了新的指导，媒介融合的目标不是利用新技术完成新功能的简单嫁接，而是在受

[1] Gibson，J. J. ，The ecological approach to visual perception ， Hillsdale，NJ: Lawrence Erlbaum，1986.
[2] 喻国明,赵睿.媒体可供性视角下“四全媒体”产业格局与增长空间[J].学术界,2019(07):37-44.

众与媒介相互作用的过程中诞生新的“可供性”。面对方兴未艾的AI技术，出版产业如何在传播实践中探索新的“媒体可供性”，为受众提供丰富高效的媒介体验，是其激发新增长空间的战略方向。

根据艾瑞网《2021年中国网络音频产业研究报告》显示中国网络音频产业规模预计在2022年达到229亿元，有声读物仍然处于欣欣向荣的发展态势，并且AI有声书迅速崛起已与真人有声书呈现并驾齐驱的趋势。AI有声书即通过语音合成技术（TTS）将文本信息转换为具有人类情感和语言特征的音频信号，进而以语音识别技术（ASR）完成口语式人机交互。AI有声书在制作环节以全自动化的长文本理解、语音合成控制、自动化后期替代了真人有声书需要人工完成的“画本”、试音、录音、审听、后期等环节，甚至可以通过语音交互与有声书实现“对话”。

有声书的历史远早于印刷书籍，在印刷机器发明之前以“口-耳传播”的“吟游诗人”对平民阶层的文化生活产生深远影响，我国的“评书”艺术直接影响了通俗文学发展。1877年爱迪生发明留声机标志有声书进入“口—机器—耳传播”阶段，最初有声书广泛应用于以盲人、儿童为受众的公益项目，随着信息技术进步，有声书的发展经历了从盒式磁带到CD再到数字格式的三个阶段，如今已成为出版产业的重要组成部分。2011年至2013年间蜻蜓FM、喜马拉雅、懒人听书先后上线手机客户端，我国的有声书产业进入移动互联网阶段。在产业层面上形成综合在线音频平台与垂直听书类平台差异化竞争，电子书平台听书功能、知识付费平台听书功能、阅读类微信公众号听书功能和谐共存，在此局面下传统出版社势微。

在AI语音技术的催化下，有声书开始了“机器生产”-“机器传播”-“用户体验”-“人机交互”的循环传播模式，下面以“媒体可供性”视角简要分析该模式（详见图1）。AI语音技术主要形塑了有声书的生产环节，更高效的内容生产力带来的新可供性表现为：有声书数月的制作周期被缩短为实时合成；前期的“画本”制作、中期的配音，后期的审听包装可以交由机器实现全流程的自动化；技术生产要素为主导的边际成本趋零；通过迭代技术与离线合成实现AI有声书

的精品化。生产可供性的改善使内容供给端更为重视 AI 有声书的应用，如数字出版“起点”、问答社区“知乎”、新闻媒体“财新”等平台全方位的引用 AI 语音技术。在融媒体与声音经济的背景下，这些原本擅长图文内容的平台通过引入技术以较低的成本快速搭建起自身的音频传播能力。在读者体验端 AI 有声书提供了更易得且便宜的有声书、个性化的音色选择、多元的阅读场景、无障碍的语音交互。AI 语音技术的普及使文字到音频可以一键转换，视觉与听觉结合构成的多媒体阅读为读者带来更好的体验，例如，作家、诗人、短视频创作者“叙矣”（抖音号：xuyi59）把 AI 语音技术合成的诗作上传在抖音平台，它们的点赞量是他普通图文作品的 200 倍左右。目前，AI 有声书依然存在情感、停顿、重音、语调、语速不够自然真实的问题，但创造新媒介的技术对内容生产与传播的改造是一个熵增过程，即由简单向复杂发展且不可逆，随着人工智能技术进步 AI 有声书有望成为未来有声书的主流形式，出版社应该把握住 AI 有声书快速发展的机遇期，探索新的业务增长点。

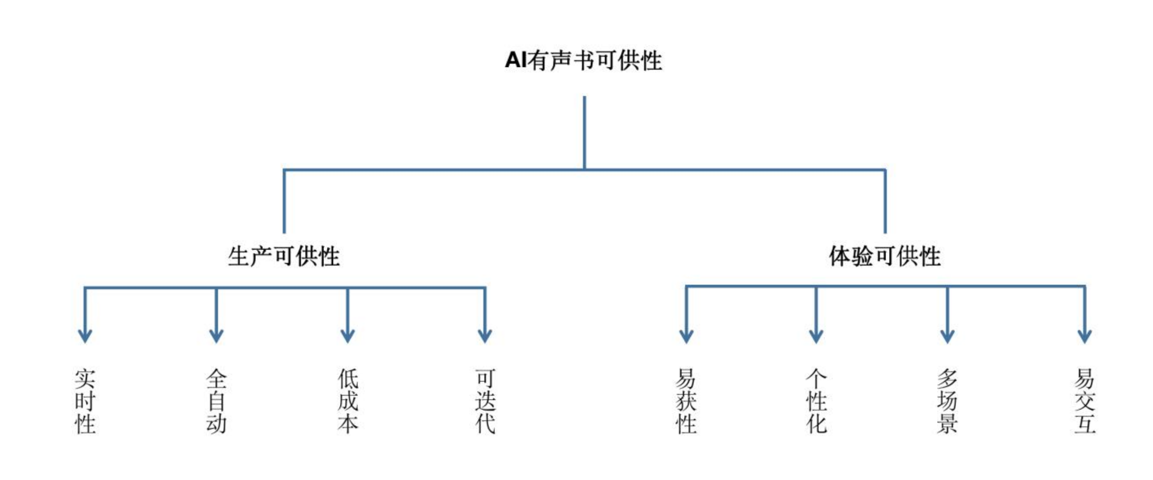


图 1 AI 有声书可供性

二、出版社发展 AI 有声书面临的问题

2.1 现行著作权法不完全适用，增加产业发展不确定性

著作权法是印刷技术的产物，其产生和发展一直与技术的进步紧密相连，表现为著作权的客体不断增加和对作品的利用方式不断丰富。[3]基于 AI 语音技术对作品的二次创作在我国现行著作权法下仍属于模糊范畴。AI 语音合成是否侵犯作品的表演权或复制权，实时和非实时语音合成在法律界定上有什么不同，AI 语音技术完成的二次创作音频是否拥有著作权，AI 语音技术模仿某一真实人声创作的作品著作权归属？以上问题仍未形成法律共识。由于缺少清晰的法律界定，出版社在制作和运营 AI 有声书的过程中更容易产生经济纠纷，滞后的著作权体系为这项业务增添了巨大的风险。

2.2 IP 热潮破坏原有生态，全版权库建设后继乏力

“内容产业”的核心是 IP，围绕一个 IP 可以进行影视剧改编、游戏开发、音乐创作、二次元创作、衍生品开发等多种文化产业运作，结果产生更大的经济效益。[4]音频作品作为一种以声音为表现形式的内容形态已经被纳为 IP 生态产业的重要一环，在此背景下出版社却越来越难获得作者的全版权授权。主要原因是出版社缺乏全版权运营的能力，或是优质 IP 在出版前已授出其他权利。对于出版社来说优质 IP 是核心资源也是稀缺资源，围绕它可以开发出多项经济收益。但当前出版社在签新书时往往只能获得图书出版授权，缺乏全版权库建设能力对于出版社开展 AI 有声书业务犹如无米之炊。

2.3 新业务亟需新技术人才，管理落后加重恶性循环

新业务的开展需要持续投入大量的专业人力资源，虽然技术进步提升了工作效率，但新技术的应用也对人力素质提出更高的要求。传统出版行业属于知识密集型产业，聚集了大量优秀的编校人才，但在数字信息时代技术与运营的重要性日益突显。企业文化缺乏创新性、人力资源成本上升、人力管理缺乏激励性等问题使传统出版社既难以吸引新型人才，又造成原有优秀人才流失，长此以往恶性循环。智联招聘发布的 2021 年第四季度《中国企业招聘薪酬报告》显示，出版行业平均薪酬为 9073 元/月，在 48 个行业类别中排名 35 位，整体处于中下游水

[3] 李明德，许超. 著作权法[B]. 北京：法律出版社，2009.
[4] 王志刚，李阳冉. 知识管理视角下网络文学 IP 生态体系重构[J]. 编辑之友，2021（05）：40-45.

平。在人力资源高度市场化的背景下，工资正成为职业选择的决定性因素，出版社对于人才的吸引力正逐渐下降。而 AI 有声书的制作与运营需要有技术基础和运营经验的人才，此类人才的招聘与培训需要出版社持续投入大量成本。

2.4 融合出版增加内容生产传播复杂度，全平台运营难管理

融合出版要求出版社具备“一次生产、多次加工、多功能服务、多载体（渠道）传播”的能力，对应的每一环节都需要投入具备技术或经验的专业性人力资源。其中多功能服务与多载体（渠道）传播主要指“全平台运营能力”，AI 有声书在完成制作后便进入运营阶段。与传统的图书发行不同，虚拟存在的 AI 读物属于内容服务型产品，内容品质与服务体验共同决定读者的阅读体验。并且它的传播能力与复制（印刷）数量无关，而与传播的平台、媒介相关，全平台运营的内容往往可以获得更好的传播效果。与全平台运营能力相对应的是更大的运营团队，每增加一个传播平台或媒介时运营人力的投入也需要倍数型增长。AI 有声书仅是众多内容形式中的一种，中小型出版社无力对其全平台运营。

2.5 超级科技平台强化垄断优势，中小出版社被剥夺话语权

2021 年 10 月，《中华人民共和国反垄断法（修正草案）》进行初次审议，现已实施 13 年的《反垄断法》迎来首次修正，释放出打击平台垄断的强监管信号。得益于我国对新业态、新模式秉持包容审慎监管的态度，互联网与人工智能等产业迅猛发展，AI 有声书在内容发行与技术支持等产业链环节已形成数个具有垄断优势的平台。超级平台市场势力过于强大，已严重危及市场公平竞争和技术创新，对数字平台从立(修)法到执法加强监管，已成全球共识。[5]数量众多的中小型出版社和图书公司正在产业链中受到强势平台的倾轧，在产业链上游“阅文”与“晋江”两个平台把控住 IP 的输出，AI 语音技术服务方面“科大讯飞”基本处于一家独大地位，而有声读物数字发行平台也基本处于“BAT”等科技巨头的势力范围，在超级平台面前中小型出版社和图书公司基本没有话语权。

2.6 科技巨头蔓延产业链上下游，产业显现马太效应

[5] 孙晋. 数字平台的反垄断监管[J]. 中国社会科学，2021（05）：101-127+206-207.

融合出版改变了内容生产，加快了技术融合，丰富了传播渠道，增加了服务类型，导致整个产业链的协作分工日益复杂化，处于行业中游的出版社将面临更高难度的产业协作问题。AI 有声书产业协作要求产业链上下游协同、优势互补，出版社与科技平台的传统合作模式基本上是出版社提供内容，科技公司提供技术支持，平台提供流量。但随着科技平台推行内容生态战略，其业务开始向上下游拓展，试图掌控整个产业链全流程以获取更大的经济利润。例如，“腾讯”通过在内容生态上的布局，已经获得 AI 有声书从 IP 到制作再到发行的全产业链能力。AI 有声书产业在发展初期已出现产业链下的资源马太效应，出版社正在产业协作的过程中被边缘化。

三、出版社优化 AI 有声书产品的路径

“CHESS 战略”是“产业融合理论”的经典模型，阐释了企业为实现融合发展需要采取的措施。“CHESS”中的“C”代表创新性集成，“H”代表横向组织结构，“E”代表产业规范标准的设定，“S”代表规模经济与范围经济，“S”代表系统化聚焦流程。基于“CHESS 战略”构建出版社融合发展 AI 有声书产业模型（详见图 2）并具体阐述路径策略，有助于出版产业与高新产业在技术与制度创新的基础上交叉、渗透、重组，形成新型内容产业形态。

3.1 开发迭代化：阶段式产品优化，提升效果与效率

出版社的内容生产节奏相比新兴媒体较慢，主要生产深度阅读的精品内容，一本图书从选题到发行短则数个月长则数年，而且版本更新以年为单位甚至不会更新版本。但信息时代瞬息万变，内容具有很强的时效性，读者的喜好与需求倒逼内容持续优化迭代，技术更新也要求内容形式和传播方式需要不断提升迭代。对于 AI 有声书，在出版社缺少内容制作经验，AI 语音技术仍未完全成熟的条件下，需要出版社通过内容产品迭代的方式积累制作经验、适配技术升级，以制作出不断满足读者新需求的内容产品。相较于传统瀑布模型开发方式中以完成完整

的系统项目为目标，迭代化方式将整个项目目标按照逻辑结构划分为一个个易于执行的小任务。通过迭代化的开发方式可以快速将 AI 有声书推向市场，然后根据用户的需求反馈不断迭代系统，增加新的功能模块，实现高质量、高效率的 AI 有声书。例如，中信出版集团 2017 年开始打造的“中信书院”最初以数字阅读为主，经过多次迭代并引入科大讯飞的 AI 语音技术，现在已经发展成为包括文字、音频、视频的全形态、系统性多媒体知识服务平台，大量读者选择付费阅读由 AI 语音合成技术生成的有声读物。

3.2 经营差异化：利用长尾效应与头部精品差异化竞争

目前真人有声书依然主导市场，以有声书市场份额占比最大的平台喜马拉雅为例，虽然其上线了大量由 AI 语音技术生成的有声书，但位于排行榜前列的头部作品均为知名主播录制，且配音演员作为卖点之一被体现在标题上。此外，在知识付费和垂直内容领域大 V 的作用难以取代，真人有声书情感细腻的听觉体验和更为自由的二次创作对于 AI 语音技术来说短期难以实现。因此，AI 有声书的商业路径需要采取差异化策略，利用其低成本、短周期、快速大量合成的优势主攻中腰部和尾部作品，这一策略恰好符合长尾效应（The Long Tail Effect）。主打免费音频的番茄畅听，作为网络音频赛道的新入局者通过积极引入 AI 语音技术获得竞争优势，在其内容分类中“真人朗读”与“AI 朗读”作为重要的分类标签并列排布，且“AI 朗读”在有声书数量、收听人数、评分三个重要指标上均在接近“真人朗读”。出版社应该将储备的 IP 资源进行划分，头部 IP 自主或授权第三方录制成真人精品有声书，而中腰部 IP 低成本、大批量生成 AI 有声书，以实现帕累托最优（Pareto Optimality）。

3.3 业务平台化：构建内容分发平台，推动运营全面化

在强调信息整合与价值分配的 Web3.0 时代，出版社迫切需要构建自主的内容分发平台，以掌握主动权降低对超级平台的依赖。目前出版社构建平台主要有两条路径：一是借助社交媒体的流量入口开发小程序，二是依靠内容、服务、品牌构建网站（App）平台。第一条路径具有推广成本低、开发门槛低、无须用户

下载、操作体验佳、对网页展示兼容性强等优点，[6]但在借助社交媒体流量的同时也会加深对其依赖，且具有入口深、功能简单、不稳定、内容传播效果差等缺点。阿拉丁研究院发布的《2021 年小程序互联网发展白皮书》显示，全网小程序数量已超 700 万，其中微信小程序开发者突破 300 万，小程序 DAU 已超 4.5 亿；日均使用次数同比增长 32%，活跃小程序则增长 41%。[7]其中人民文学出版社、中华书局、高等教育出版社等出版社已上线小程序，整体来看小程序较适合出版社优化服务和促进内容付费。第二条路径对于中小型出版社难度较高，并不适合所有出版社，需要其具有提供不可替代服务或商品的能力。但其优势也是明显的，网站（App）平台的建立将加强其护城河，例如高等教育出版社旗下的“中国大学 MOOC”是一个成功案例，它通过应用 AI 语音识别技术为音视频内容快速生成字幕。但是，构建自主内容分发平台并不是要舍弃由互联网巨头控制的平台，相反出版社要加强 AI 有声书的全平台运营能力，这既有利于增强传播效能，也有助于遏制超级平台的垄断。

3.4 IP 产品化：以产品思维开展营销，与分发平台和谐共生

“产品经理”一词近年来频频见诸于出版行业。产品经理引入出版产业是“产业融合”发展的产物，也是内在运作机理的必然要求。[8]虽然图书营销专员与图书产品经理之间分工不同，但畅销书的成功需要图书产品经理从选题开发到售后服务的整个闭环中充分考虑营销的影响。AI 有声书的产品创作过程同样要充分考虑到营销环节，既是为了最大化 IP 的商业价值也是为了增进 IP 的持续影响力。微信、抖音、喜马拉雅等大平台为 AI 有声书的传播提供了更具效能的渠道，提供 IP 内容的出版社与提供流量的平台是一种和谐共生的关系。《2021 中国网络视听发展研究报告》显示，喜马拉雅的用户渗透率达到 67.1%，牢牢占据着在线音频行业第一梯队的地位，全端平均月活跃用户为 2.68 亿。因此，出版社发展 AI 有声书同真人有声书一样需要增加在喜马拉雅 FM 等网络音频平台的内容分发，并且这与搭建自主的内容平台不冲突。“产业融合”不仅从微观上改变了产业的

[6] 刘越. 基于微信小程序的应用在医院信息系统中的研究[J]. 世界最新医学信息文摘(连续型电子期刊)，2020（92）：222-223.
[7] 阿拉丁研究院. 2021 年小程序互联网发展白皮书[EB/OL].（2022-01-17）[2022-06-11]. https://aldzs.com/viewpointarticle?id=16175.
[8] 王建，付小艳. 数字出版编辑思维转型：从编辑走向产品经理[J]. 中国出版，2019（04）：35-38.

市场结构和产业绩效，而且从宏观上改变了一个国家的产业结构和经济增长方式。[9]内容出版与 AI 技术“产业融合”既可以减少企业成本，也是传统产业创新的重要方式和手段，有利于出版产业结构转换升级，提高国家文化竞争力。

3.5 技术服务化：与技术企业合作共赢，支持技术产品服务化

“产业融合”发生的基础是技术进步和放松管制，阿里云全球技术服务总经理李津提出“从技术走向产品，再从产品走向服务是所有技术企业的必由之路。”产业互联网新经济形态正在形成，各个垂直行业的产业链被其重塑和改造，出版行业应该主动利用由信息技术与互联网平台提供的技术服务产品，提升技术生产要素在其内容生产种的比例，通过技术创新提升出版社生产力。当前技术产品服务化正呈现出技术平台化、云端化、标准化、基础化，服务集成化、多元化、个性化、可协同、跨行业的特点。“产业融合”改变了企业之间的竞争合作关系，科大讯飞等以 AI 语音技术为核心的开放平台通过提供技术服务方案与包括出版行业在内的各行各业实现合作共赢。应用于 AI 有声书的技术需要具备音频采样与编码、语音识别数据库匹配、语音转文本、长文本理解、情感语音合成、自动化后期的能力，以上每一种能力都需要拥有先进技术储备。例如，情感语音合成需将文本情感与声音情感匹配，并且为声音增加符合人类语言习惯的停顿、重音、语调、语速效果。优秀的合成语音可以超出人耳对于声音情感的辨别能力，目前情感语音合成依然是行业难题，商用 AI 语音合成技术的情感区分度基本处于 8 种及以下。由此可见 AI 语音技术存在极高的技术门槛，出版社与技术企业合作共赢，支持技术产品服务化成为必然选择。

[9] 张建刚，王新华，段治平. 产业融合理论研究述评[J]. 山东科技大学学报（社会科学版），2010，12（01）：73-78.

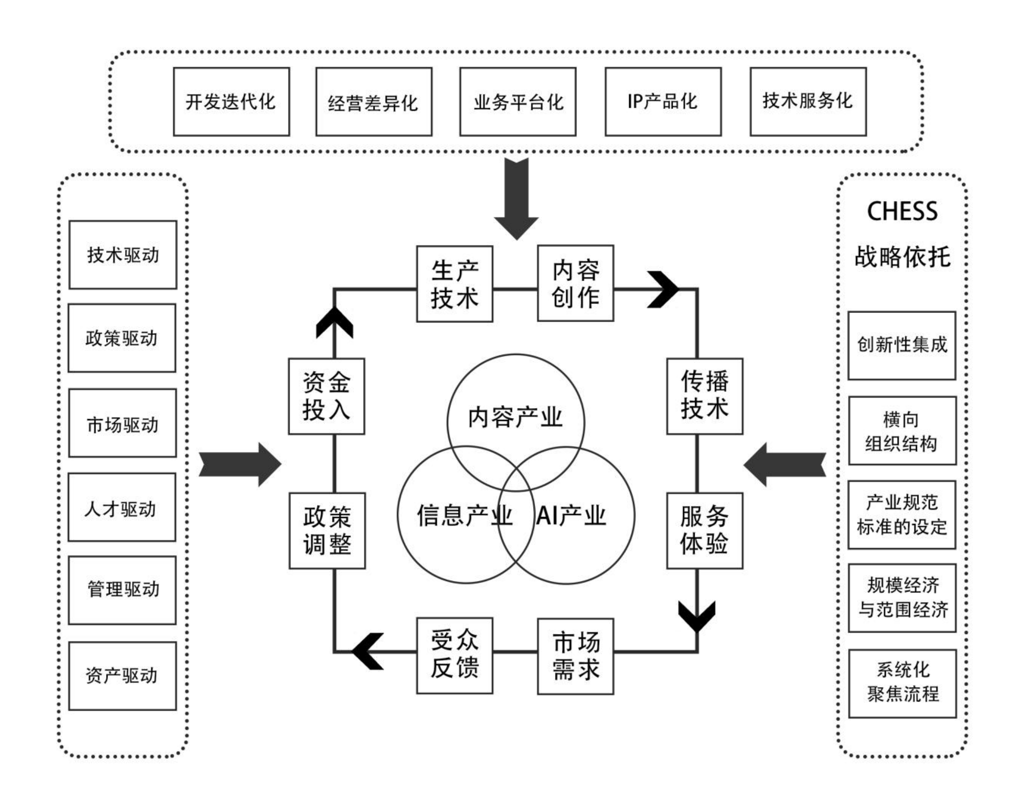


图 2 出版社 AI 有声书“产业融合”发展模型图